Chapter

8

Some Advanced Topics on Optimization

8.1 Stochastic Optimization

Global optimization algorithms are categorized into two major classes: Deterministic and stochastic (Dixon *et al.* 1975). In some chemical engineering optimization problems, the model cannot be entirely specified because, it depends on quantities that are random or probabilistic at the time of formulation. For instance, uncertainty governs the prices of fuels, the availability of electricity, and the demand for chemicals in the market [N. V. Sahinidis, (2004)]. These ever-changing market conditions require a high flexibility for chemical process industries under various product specifications and various feedstocks. Furthermore, the properties of processes change themselves during process operation, e.g., tray efficiencies and fouling of the equipment (heat exchanger, catalytic reactor), which leads to a decrease of product quality with the same operating conditions [Henrion *et al.* (2001)]. For deterministic optimization methods, the expected values of uncertain variables are generally used. In real life application, the values of uncertain variables deviate from their expected values. When we are introducing a stochastic element within deterministic algorithms, the deterministic assurance that the global optimum can be determined is relaxed into a confidence measure. Stochastic methods can be utilized to evaluate the probability of obtaining the global minimum. Stochastic ideas are generally used to development the stopping criteria, or for the approximation of the regions of attraction as utilized by some methods [Arora *et al.* 1995].

8.1.1 Uncertainties in process industries

Uncertainty is a part of real systems; in reality, every phenomenon has a certain degree of uncertainty. During their operation, chemical processes are frequently faced with uncertain conditions. From the perspective of chemical process operation, there exist two types of uncertainties. These uncertainties can occur due to variation either in external parameters, such as fluctuation of inlet flow, temperature, pressure, composition of the feedstock, supply of utilities and other environment

parameters, or in internal process parameters for example heat transfer coefficients, tray efficiency and reaction rate constants, etc. The optimization problems become much more complex due to the existence of uncertain parameters and have considerable implications on process feasibility and other quality measures for instance controllability, safety, and environmental compliance [Hou *et al.* (2000)]. Uncertainty also arises in the planning problem. Uncertainty in planning problems can be categorized into two groups: endogenous uncertainty and exogenous uncertainty. Problems where stochastic processes are affected by decisions are said to possess endogenous uncertainty while problems where stochastic processes are independent of decisions are said to have exogenous uncertainty (e.g., yields, demands) [Jonsbraten, (1998)]. A typical example of stochastic process is discussed in the following section.

Fluidization Fluidization is a very common application in chemical process plant. In the chemical process industry, fluidized bed reactors are among those that have been utilized extensively for various applications. During fluidization, movement of the particles are highly stochastic; different particle has different residence time. Dehling *et al.* [Dehling *et al.* (1999)] investigated residence time distributions in continuous fluidized beds employing stochastic modeling. They proposed a stochastic model for fluidization operation.

To elucidate the problem, a batch fluidized bed has been considered. There is no inflow and outflow of the particles when the fluidized bed is in operation. The discretized form of a fluidized bed has shown in the Fig. 8.1(b). The movement of a single particle is considered, and the transport processes are transformed to transition probabilities between cells as shown in Fig. 8.1(b). The behavior of a particle has been reflected by the probability distribution for the particle's position as a function of time. The model is developed based on Markov chains so that the transition probability distribution of a single particle does not depend on the past history of the system.

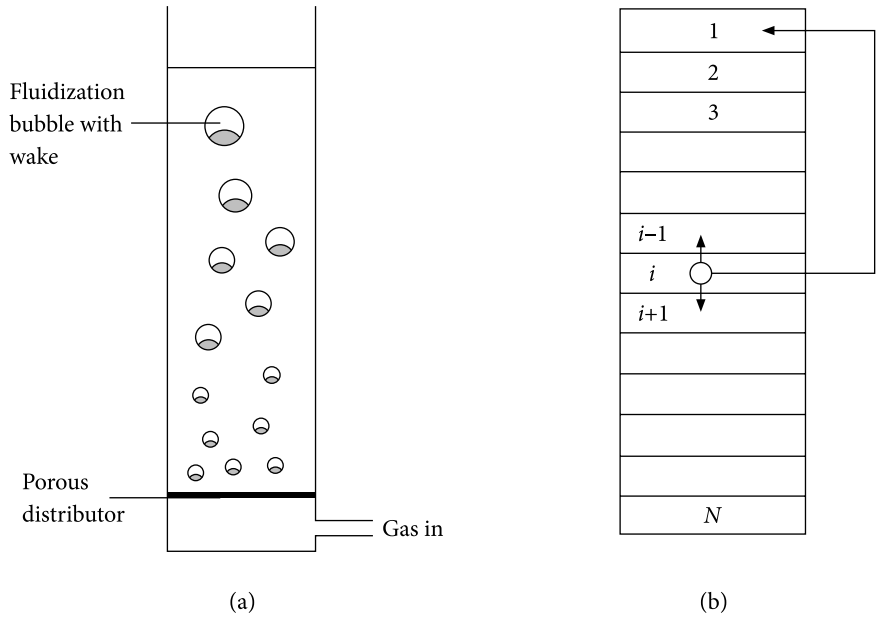


Fig. 8.1 Fluidization column

In this discrete Markov model, the fluidized bed reactor is partitioned into N horizontal cells, and the position of the particle was modeled at discrete times only. The cell numbers are given from the top toward the bottom as shown in Fig. 8.1. This model determines the probability distribution of the particle's axial position as a function of time. The probable transitions are as follows:

i. staying in the same cell (i)
ii. traveling to the next (above) cell (i –1)
iii. moving back to the previous (below) cell (i + 1)
iv. deposited at the top of the bed (cell 1) after getting caught up in a bubble wake

8.1.2 Basic concept of probability theory

Probability and statistics are concerned with events that occur by chance. Examples consist of an occurrence of events such as opening of an off-on valve, errors of measurements, concentration of contaminants in wastewater, demand of a particular petroleum product. In each event, we may have some information on the likelihood of different probable outcomes, but, we are not able to predict with any certainty the outcome of any particular trial. Probability and statistics are applied all over engineering applications. Signals and noise are analyzed using probability theory in electrical engineering. Mechanical, civil, and industrial engineers use probability and statistics to test and account for variations in materials and goods. To control and improve the chemical processes, chemical engineers use probability and statistics to evaluate experimental data. In recent times, it is an essential requirement for all engineers [DeCoursey (2003)].

Preliminaries

Probability Every incident in reality possesses a certain element of uncertainty.

A basic requirement in probability theory is random experiment: an experiment whose outcome cannot be decided in advanced. The set of all potential outcomes of an experiment is called the sample space of that experiment, and represented by S.

An event is a subset of a sample space, and said to occur if the outcome of the experiment is an element of that subset. We can define the probability in terms of the likelihood of a particular event. If an event is denoted by E, the probability of occurrence of the event E is generally represented by $P(E)$. The probability of occurrence depends on the number of trials or observations. It can be written as

$$P(E) = \lim_{n\to\infty} \frac{m}{n} \tag{8.1}$$

where m indicates the number of successful occurrence of the event E and the total number of trials is denoted by n. From Eq. (8.1) we can say that probability is a non-negative number and

$$0 \le P(E) \le 1 \tag{8.2}$$

$$P(S) = 1 \tag{8.3}$$

For any series of events $E_1, E_2 \ldots$ that are mutually exclusive, that is, events for which $E_i E_j = \phi$ when $i \neq j$ (where ϕ is the null set)

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P\left(E_i\right) \tag{8.4}$$

the probability of the event E is represented as $P(E)$.

Random variable

Consider a random experiment having sample space S. A random variable X is a function that assigns a real value to each outcome in S. An event can be defined as a probable outcome of an experiment. Let us consider that a random event is the measurement of a quantity X that has different values within the range $-\infty$ to ∞. This quantity (X) is called a random variable. Random variable is denoted by a capital letter and the particular value taken by it is represented by a lowercase letter.

For any real number x, the distribution function F of the random variable X can be defined by

$$F\left(x\right) = P\left\{X \leq x\right\} = P\left\{X \in \left(-\infty, x\right)\right\} \tag{8.5}$$

we shall denote $1 - F(x)$ by $\bar{F}\left(x\right)$, and so

$$\bar{F}\left(x\right) = P\left\{X > x\right\} \tag{8.6}$$

There are two types of random variables: (i) discrete and (ii) continuous. When the random variable is permitted to use only discrete values $x_1, x_2, \ldots, x_n$, it is called a discrete random variable. A random variable X is called discrete when its set of possible values is countable. For discrete random variables,

$$F\left(x\right) = \sum_{y \leq x} P\left\{X = y\right\} \tag{8.7}$$

In contrast, when the random variable is allowed to use any real value within a specified range, it is called a continuous random variable. A random variable is called continuous if there exist a function $f(x)$, called the probability density function, such that

$$P\left\{X \text{ is in } B\right\} = \int_B f\left(x\right) dx \tag{8.8}$$

for every set B. Since, $F\left(x\right) = \int_{-\infty}^{x} f\left(x\right) dx$, it follows that

$$f\left(x\right) = \frac{d}{dx} F\left(x\right) \tag{8.9}$$

For example, the number of worker present in any day, number of tray in a distillation column are discrete random variables, while the inlet composition of a distillation column can be considered as a continuous random variable.

Expected value

The expected value or mean of the random variable X is represented by $E(X)$

$$E(X) = \int_{-\infty}^{\infty} x dF(x) \tag{8.10}$$

$$= \begin{cases} \int_{-\infty}^{\infty} x f(x)\, dx & \text{if } X \text{ is continuous} \\ \sum_{x} x P\{X = x\} & \text{if } X \text{ is discrete} \end{cases} \tag{8.11}$$

provide the above integral exists.

Equation (8.11) also defines the expected value of any function of X, say $h(X)$. Since, $h(X)$ itself is a random variable, it follows from Eq. (8.11) that

$$E\left[h(X)\right] = \int_{-\infty}^{\infty} x dF_h(x) \tag{8.12}$$

where F_h is the distribution function of $h(X)$. Though, we can show that this is identical to $\int_{-\infty}^{\infty} h(x) dF(x)$. That is

$$E\left[h(X)\right] = \int_{-\infty}^{\infty} h(x) dF(x) \tag{8.13}$$

The variance of the random variable X is defined by

$$\begin{aligned} \text{Var}\, X &= E\left[\left(X - E[X]\right)^2\right] \\ &= E\left[X^2\right] - E^2[X] \end{aligned} \tag{8.14}$$

Two jointly distributed random variables X and Y are said to be uncorrelated if their covariance, define by

$$\begin{aligned} \text{Cov}(X,Y) &= E\left[(X - EX)(Y - EY)\right] \\ &= E[XY] - E[X]E[Y] \end{aligned} \tag{8.15}$$

is zero. It follows that independent random variables are uncorrelated. However, the converse need not be true.

An important property of expectations is that the expectation of a sum of random variables is equal to the sum of the expectations.

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] \tag{8.16}$$

The corresponding property for variances is that

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum\sum_{i<j} \mathrm{Cov}(X_i X_j) \tag{8.17}$$

Standard Deviation

The expected value or mean is a measurement of the central tendency, which indicates the location of the distribution on some coordinate axis. The variability of the random variable is measured by a quantity that is called the standard deviation.

The variance or mean-square deviation of a random variable X can be written as

$$\begin{aligned}\sigma_X^2 &= \mathrm{Var}(X) = E\left[(X-\mu_X)^2\right] \\ &= E\left[X^2 - 2X\mu_X + \mu_X^2\right] \\ &= E(X^2) - 2\mu_X E(X) + E(\mu_X^2) \\ &= E(X^2) - \mu_X^2\end{aligned} \tag{8.18}$$

the standard deviation is defined as

$$\sigma_X = +\sqrt{\mathrm{Var}(X)} = \sqrt{E(X^2 - \mu_X^2)} \tag{8.19}$$

Example 8.1

Number of worker present in a chemical industry and their probability is given in Table 8.1

Table 8.1 Number of worker present with probability

x_i	0	1	2	3	4	5	6	7
$P_x(x_i)$	0.01	0.12	0.22	0.25	0.15	0.13	0.07	0.05

Find the mean and standard deviation of x

Solution

Using the data from the above table,

$$\bar{x} = \sum_{i=0}^{7} x_i P_x(x_i) \tag{8.20}$$

$$= 0(0.01) + 1(0.12) + 2(0.22) + 3(0.25) + 4(0.15) + 5(0.13) + 6(0.07) + 7(0.05)$$

$$= 0 + 0.12 + 0.22 + 0.75 + 0.60 + 0.65 + 0.42 + 0.35$$

$$= 3.33$$

and

$$\sum_{i=0}^{7} x_i^2 P_x(x_i) = 0(0.01) + 1(0.12) + 4(0.22) + 9(0.25) + 16(0.15) + 25(0.13) + 36(0.07) + 49(0.05) \tag{8.21}$$

$$= 0 + 0.12 + 0.88 + 2.25 + 2.40 + 3.25 + 2.52 + 2.45$$

$$= 13.87$$

Standard deviation $= 13.87 - (3.33)^2$

$$= 2.7811$$

Stochastic Process A stochastic process is a collection of random variables that can be represented as $\underline{X} = \{X(t), t \in T\}$. Where $X(t)$ is a random variable for each t in the index set T. Most often, t is interpreted as time and describe $X(t)$ the state of the process at time t. If the index set T is a countable set, we call $\underline{X}$ a discrete-time stochastic process, and if T is a continuum, we call this process a continuous-time process.

8.1.3 Stochastic linear programming

A simple stochastic linear programming problem is represented as

$$\text{Minimize } f(X) = C^T X = \sum_{j=1}^{n} c_j x_j \tag{8.22a}$$

subject to

$$A_i^T X = \sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad i = 1,,2.....m \tag{8.22b}$$

$$x_j \ge 0, \qquad j = 1,,2.....\text{n} \tag{8.22c}$$

where c_j, a_{ij}, and b_i are random variables (for simplicity we are assuming that the decision variables x_j are deterministic) with known probability distributions. Numerous methods are available for solving the problem given by Eqs (8.22a)–(8.22c). Chance-constrained programming technique has been considered in this section.

Charnes and Cooper [Charnes and Cooper, (1959)] originally developed the method chance-constrained programming technique. It is clear from the name that the method can be employed to solve problems concerning chance constraints that is, constraints may be violated with finite probability. The stochastic programming problem can be represented as follows:

$$\text{Minimize } f(X) = \sum_{j=1}^{n} c_j x_j \tag{8.23a}$$

subject to

$$P\left[\sum_{j=1}^{n} a_{ij} x_j \le b_i\right] \ge p_i, \quad i = 1,,2.....m \tag{8.23b}$$

$$x_j \ge 0, \quad j = 1,,2.....n \tag{8.23c}$$

where p_i are specified probabilities and c_j, a_{ij}, b_i are random variables. Notice that Eq. (8.23b) indicate that $\sum_{j=1}^{n} a_{ij} x_j \le b_i$ (the ith constraint) has to be satisfied with a probability of at least p_i where $0 \le p_i \le 1$. To make it simple, we consider that the design variables x_j are deterministic and c_j, a_{ij}, and b_i are random variables. We also consider that all the random variables are normally distributed with known mean and standard deviations.

Since, c_j are normally distributed random variables, the objective function $f(X)$ will also be a normally distributed random variable. The value of mean and variance of the function f are as follows

$$\overline{f} = \sum_{j=1}^{n} \overline{c}_j x_j \tag{8.24}$$

$$\text{Var}(f) = X^T V X \tag{8.25}$$

where $\overline{c}_j$ denotes mean value of c_j and the covariance matrix of c_j is defined by the matrix V (Eq. (8.26)).

$$V = \begin{bmatrix} \mathrm{Var}(c_1) & \mathrm{Cov}(c_1,c_2) & \cdots & \mathrm{Cov}(c_1,c_n) \\ \mathrm{Cov}(c_2,c_1) & \mathrm{Var}(c_2) & \cdots & \mathrm{Cov}(c_2,c_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(c_n,c_1) & \mathrm{Cov}(c_n,c_2) & \cdots & \mathrm{Var}(c_n) \end{bmatrix} \tag{8.26}$$

where Var(c_j) representing the variance of c_j and covariance between c_i and c_j are represented by Cov(c_i, c_j). The formulated new deterministic objective function for minimization is given by

$$F(X) = k_1\overline{f} + k_2\sqrt{\mathrm{Var}(f)} \tag{8.27}$$

where the constants k_1 and k_2 are nonnegative. The values of these constants signify the relative importance of the mean of the function, $\overline{f}$ and standard deviation of f during minimization. When we minimize the expected value of f without considering the standard deviation of f, the value of $k_2 = 0$. In contrast, $k_1 = 0$ signifies that we are paying attention to minimize the variability of f about its mean value without caring for the mean value of f. When equal importance is given to both the minimization of the mean and the standard deviation of f, we have to consider $k_1 = k_2 = 1$. Equation (8.27) shows that the new objective function is a nonlinear function of X.

The constraints of Eq. (8.23b) can be represented as

$$P[h_i \le 0] \ge p_i, \quad i = 1,2.....m \tag{8.28}$$

where the new random variable h_i is given as

$$h_i = \sum_{j=1}^{n} a_{ij}x_j - b_j = \sum_{k=1}^{n+1} q_{ik}y_k \tag{8.29}$$

where

$$q_{ik} = a_{ik}, \quad k = 1,2,.....,n \qquad q_{i,n+1} = b_i \tag{8.30}$$

$$y_k = x_k, \quad k = 1,2,.....,n \qquad y_{n+1} = -1 \tag{8.31}$$

Here, for convenience we introduce the constant y_{n+1}. The h_i will follow normal distribution as it is given by a linear combination of the normally distributed random variables q_{ik},. The mean and the variance of h_i are given below

$$\overline{h_i} = \sum_{k=1}^{n+1} \overline{q}_{ik}y_{ik} = \sum_{j=1}^{n} \overline{a}_{ij}x_j - \overline{b}_i \tag{8.32}$$

$$\mathrm{Var}\left(h_i\right) = Y^T V_i Y \tag{8.33}$$

where

$$Y = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n+1} \end{Bmatrix} \tag{8.34}$$

$$V_i = \begin{bmatrix} \mathrm{Var}\left(q_{i1}\right) & \mathrm{Cov}\left(q_{i1}, \mathrm{q}_{i2}\right) & \cdots & \mathrm{Cov}\left(q_{i1}, \mathrm{q}_{i,n+1}\right) \\ \mathrm{Cov}\left(q_{i2}, \mathrm{q}_{i1}\right) & \mathrm{Var}\left(q_{i2}\right) & \cdots & \mathrm{Cov}\left(q_{i2}, \mathrm{q}_{i,n+1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(q_{i,n+1}, \mathrm{q}_{i1}\right) & \mathrm{Cov}\left(q_{i,n+1}, \mathrm{q}_{i2}\right) & \cdots & \mathrm{Var}\left(q_{i,n+1}\right) \end{bmatrix} \tag{8.35}$$

More explicitly, we can write this as

$$\begin{aligned} \mathrm{Var}\left(h_i\right) &= \sum_{k=1}^{n+1}\left[y_k^2\ \mathrm{Var}\left(q_{ik}\right) + 2\sum_{l=k+1}^{n+1} y_k y_l\ \mathrm{Cov}\left(q_{ik}, q_{il}\right)\right] \\ &= \sum_{k=1}^{n}\left[y_k^2\ \mathrm{Var}\left(q_{ik}\right) + 2\sum_{l=k+1}^{n} y_k y_l\ \mathrm{Cov}\left(q_{ik}, q_{il}\right)\right] \\ &\quad + y_{n+1}^2 \mathrm{Var}\left(q_{i,\mathrm{n}+1}\right) + 2 y_{n+1}^2 \mathrm{Cov}\left(q_{i,\mathrm{n}+1}, q_{i,\mathrm{n}+1}\right) \\ &\quad + \sum_{k=1}^{n}\left[2 y_k y_{n+1} \mathrm{Cov}\left(q_{ik}, q_{i,\mathrm{n}+1}\right)\right] \\ &= \sum_{k=1}^{n}\left[x_k^2\ \mathrm{Var}\left(a_{ik}\right) + 2\sum_{l=k+1}^{n} x_k x_l\ \mathrm{Cov}\left(a_{ik}, a_{il}\right)\right] \\ &\quad + \mathrm{Var}\left(b_i\right) - 2\sum_{k=1}^{n} x_k \mathrm{Cov}\left(a_{ik}, \mathrm{b}_i\right) \end{aligned} \tag{8.36}$$

Therefore, the constraints in Eq. (8.28) can be rewritten as

$$P\left[\frac{h_i-\bar{h}_i}{\sqrt{\mathrm{Var}(h_i)}} \le \frac{-\bar{h}_i}{\sqrt{\mathrm{Var}(h_i)}}\right] \ge p_i, \qquad i=1,,2.....m \tag{8.37}$$

where $\left[\left(h_i-\bar{h}_i\right)\right]\Big/\sqrt{\mathrm{Var}(h_i)}$ signifies a standard normal variable which have a mean value of zero and a variance of one. Therefore, if s_i represents the value of the standard normal variable at which $f(s_i) = p_i$ the constraints of Eq. (8.37) can be written as

$$\phi\left(\frac{-\bar{h}_i}{\sqrt{\mathrm{Var}(h_i)}}\right) \ge \phi(s_i), \quad i=1,,2.....m \tag{8.38}$$

The following deterministic nonlinear inequalities should be satisfied to satisfy the inequalities given in Eq. (8.38):

$$\frac{-\bar{h}_i}{\sqrt{\mathrm{Var}(h_i)}} \ge s_i \qquad i=1,2.....m \tag{8.39a}$$

or

$$\bar{h}_i + s_i\sqrt{\mathrm{Var}(h_i)} \le 0, \quad i=1,,2.....m \tag{8.39b}$$

So, the stochastic linear programming problem of Eqs (8.23a) to (8.23c) can be written as an equivalent deterministic nonlinear programming problem as

$$\text{Minimize } F(X) = k_1\sum_{j=1}^{n}\bar{c}_j x_j + k_2\sqrt{X^T V X}, \quad k_1 \ge 0, \quad k_2 \ge 0 \tag{8.40a}$$

subject to

$$\bar{h}_i + s_i\sqrt{\mathrm{Var}(h_i)} \le 0, \quad i=1,,2.....m \tag{8.40b}$$

$$x_j \ge 0, \quad j=1,,2.....n \tag{8.40c}$$

A stochastic linear programming problem has been developed by Ai-Othman *et al.* [Al-Othman *et al.* (2008)] for supply chain optimization of petroleum organization under market demands and prices uncertainty.

8.1.4 Stochastic nonlinear programming

A general optimization problem need to be developed as a stochastic nonlinear programming problem, when some of the parameters within the objective function and constraints vary about their mean values [S. S. Rao, (2009)]. In this present discussion, we consider that all the random variables are independent and follow normal distribution. The standard form of a stochastic nonlinear programming problem is given as

$$\text{Find values of } X \text{ that will minimize } f(Y) \tag{8.41a}$$

subject to

$$P\left[g_j(Y) \geq 0\right] \geq p_j, \quad j = 1,,2.....m \tag{8.41b}$$

where Y represents the vector of N random variables y_1, y_2, …, y_N and it consist of the decision variables x_1, x_2, …, x_n. A special case can be obtained during the present formulation, when X is deterministic. The Eq. (8.41b) indicate that the probability of realizing $g_j(Y)$ greater than or equal to zero should be greater than or equal to the specified probability p_j. By using the change constrained programming technique, the problem given in Eqs (8.41a) and (8.41b) can be transformed into an equivalent deterministic nonlinear programming problem as follows.

Objective Function

The objective function $f(Y)$ is expanded with the mean values of y_i, $\overline{y}_i$ as

$$f(Y) = f\left(\overline{Y}\right) + \sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial y_i}\right|_{\overline{Y}}\right)\left(y_i - \overline{y}_i\right) + \text{ higher-order derivative terms} \tag{8.42}$$

We can approximate the objective function $f(Y)$ by the first two terms of Eq. (8.42), if the standard deviations of $y_i\left(\sigma_{y_i}\right)$ are small

$$f(Y) \simeq \overline{Y} - \sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial y_i}\right|_{\overline{Y}}\right)\overline{y}_i + \sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial y_i}\right|_{\overline{Y}}\right)y_i = \psi(Y) \tag{8.43}$$

The $\psi(Y)$, a linear function of Y follows the normal distribution when all $y_i (i = 1, 2, \ldots, N)$ follow the normal distribution. The mean and the variance of ψ are written as

$$\overline{\psi} = \psi\left(\overline{Y}\right) \tag{8.44}$$

$$\text{Var}(\psi) = \sigma_{\psi}^2 = \sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial y_i}\right|_{\overline{Y}}\right)^2 \sigma_{y_i}^2 \tag{8.45}$$

as all y_i are independent. For the optimization purpose, a new objective function $F(Y)$ can be formulated as

$$F(Y) = k_1\bar{\psi} + k_2\sigma_\psi \tag{8.46}$$

where $k_1 \geq 0$ and $k_2 \geq 0$, and the numerical values of k_1 and k_2 indicate the relative importance of $\bar{\psi}$ and σ_ψ during minimization. The standard deviation of ψ can be dealt in another way by minimizing $\bar{\psi}$ subject to the constraints $\sigma_\psi \leq k_3\bar{\psi}$, where k_3 is a constant, along with the other constraints.

Constraints

When some of the parameters are random, then the constraints will also be probabilistic and one may wish to have the probability that a given constraint is satisfied to be larger than a certain value. This is exactly what is mentioned in Eq. (8.41b). The constraint inequality (8.41b) can be represented as

$$\int_0^\infty f_{g_j}(g_j)\,dg_j \geq p_j \tag{8.47}$$

where $f_{g_j}(g_j)$ denotes the probability density function of the random variable g_i (as the function of some random variables is also a random variable) with a range considered as $-\infty$ to ∞. The constraint function $g_j(Y)$ has been expanded around $\bar{Y}$, the vector of mean values of the random variables by the following equation

$$g_j(Y) \simeq g_j(\bar{Y}) + \sum_{i=1}^{N}\left(\left.\frac{\partial g_j}{\partial y_i}\right|_{\bar{Y}}\right)(y_i - \bar{y}_i) \tag{8.48}$$

From the Eq. (8.48), we can get $\bar{g}_j$, the mean value and σ_{g_j}, the standard deviation of g_j as given below

$$\bar{g}_j = g_j(\bar{Y}) \tag{8.49}$$

$$\sigma_{g_j} = \left\{\sum_{i=1}^{N}\left(\left.\frac{\partial g_j}{\partial y_i}\right|_{\bar{Y}}\right)^2 \sigma_{y_i}^2\right\}^{1/2} \tag{8.50}$$

a new variable is introduced as

$$\theta = \frac{g_j - \bar{g}_j}{\sigma_{g_j}} \tag{8.51}$$

and noting that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 \tag{8.52}$$

equation (11.90) can be expressed as

$$\int_{-\left(\overline{g}_j/\sigma_{g_j}\right)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} d\theta \geq \int_{-\phi_j(p_j)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{8.53}$$

where $\phi_j\left(p_j\right)$ is the value of the standard normal variate corresponding to the probability p_j. Thus,

$$-\frac{\overline{g}_j}{\sigma_{g_j}} \leq -\phi_j\left(p_j\right)$$

or

$$-\overline{g}_j + \sigma_{g_j}\phi_j\left(p_j\right) \leq 0 \tag{8.54}$$

Equation (8.54) can be written as

$$\overline{g}_j - \phi_j\left(p_j\right)\left[\sum_{i=1}^{N}\left(\left.\frac{\partial g_j}{\partial y_i}\right|_{\overline{Y}}\right)^2\right]^{1/2} \geq 0, \quad j = 1,2,\ldots,m \tag{8.55}$$

Therefore, the optimization problem given in Eqs (8.41a) and (8.41b) can be given in its corresponding deterministic form as: Minimize $F(Y)$ as given in Eq. (8.46) subject to the m number of constraints given by Eq. (8.55).

8.2 Multi-Objective Optimization

Multi-objective optimization (MOO) involves the simultaneous optimization of more than one objective function. This is quite common in the field of Chemical Engineering and Biochemical Engineering. In most of the cases, the objective functions are defined in incomparable units, and they present some degree of conflict among them. Fettaka *et al.* [Fettaka *et al.* (2013)] have described the design of heat exchanger with optimized heat transfer area and pumping power. Mass of the pipeline (kg/m) and pumping power (HP) were used as objective variables. The units of these objective functions are different. The multi-objective optimization becomes more difficult as improving one objective will worsen another. Improvement of one objective function is not possible without deteriorating at least another one.

The general multi-objective optimization problems can be represented by Eqs (8.56a)–(8.56e). Here, both equality and inequality constrains have been considered.

$$\text{Find } X = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} \tag{8.56a}$$

$$\text{which minimizes } F(X) = \left[F_1(X), F_2(X), \cdots, F_k(X)\right]^T \tag{8.56b}$$

$$\text{subject to: } g_i(X) = 0;\ i = 1,2,\cdots,m \tag{8.56c}$$

$$h_j(X) \le 0;\ \ j = 1,2,\cdots,q \tag{8.56d}$$

$$X_i^L \le X_i \le X_i^U \tag{8.56e}$$

where k is the number of objective functions, m is the number of equality constraints and q is the number of inequality constraints. $X \in E^n$ is a vector of design variables, and $F(X) \in E^n$ is a vector of objective functions $F_i(X): E^n \to E^1$. X_i^L and X_i^U represent the lower and upper boundary of the variable X_i.

Before discussing the theories of MOO, we should have ample knowledge on Pareto optimal point and utopia point (ideal point). The multi-objective optimization problem is also called the vector minimization problem. Unlike single objective optimization where there is one best solution, MOO generates a set of optimal solutions. These solutions are called Pareto optimal solutions or non-dominated solutions. The Pareto domain refers to the set of non-dominated solutions that present trade-offs among the different objectives. A solution is said to be non-dominated if it is not worse than another solution in all its objectives and it is better with respect to at least one objective (Deb, 2001).

Pareto set:

In the 1970s, Stadler [Stadler, W., (1979)] applied the concept of Pareto optimality to the fields of science and engineering. A feasible solution X^* is called Pareto optimal while there exists no other feasible solution Y such that $f_i(Y) \le f_i(X^*)$ for $i = 1,2,\ldots,k$ with $f_i(Y) < f_i(X^*)$ for at least one j. An objective vector $f(X^*)$ is called Pareto optimal if the corresponding decision vector X^* is Pareto optimal. This can be explained easily by the following example. Consider two functions $f_1 = 3 + (x-4)^2$ and $f_2 = 16 + (x-10)^2$. The minimum values of these functions are $f_{1\min} = 3$ at $x_{1\min} = 4$ at and $f_{2\min} = 16$ at $x_{2\min} = 10$ as shown in Fig. 8.2. Therefore, all the values of x between 4 and 10 are called Pareto optimal solutions. If we plot the graph f_1 vs. f_2 (Fig. 8.3), the segment a, c, d, b represents the Pareto domain.

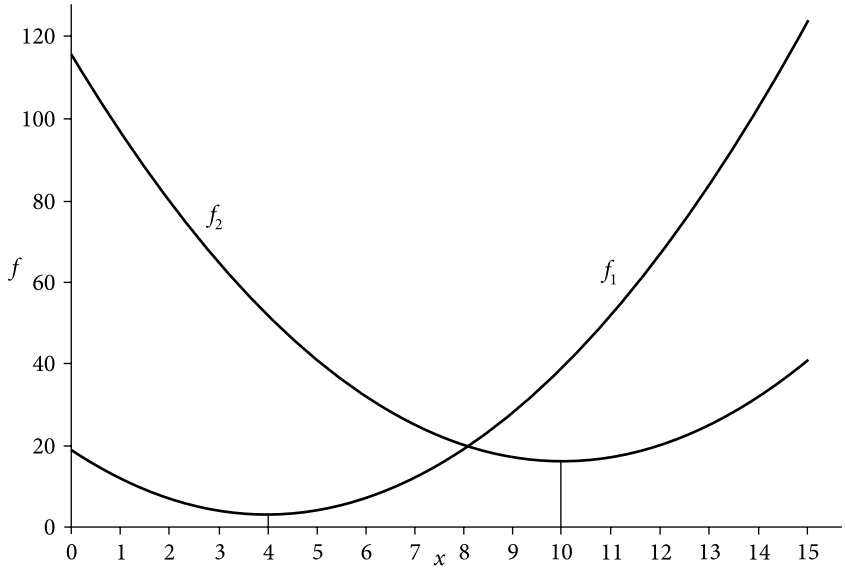


Fig. 8.2 Multi-objective optimization

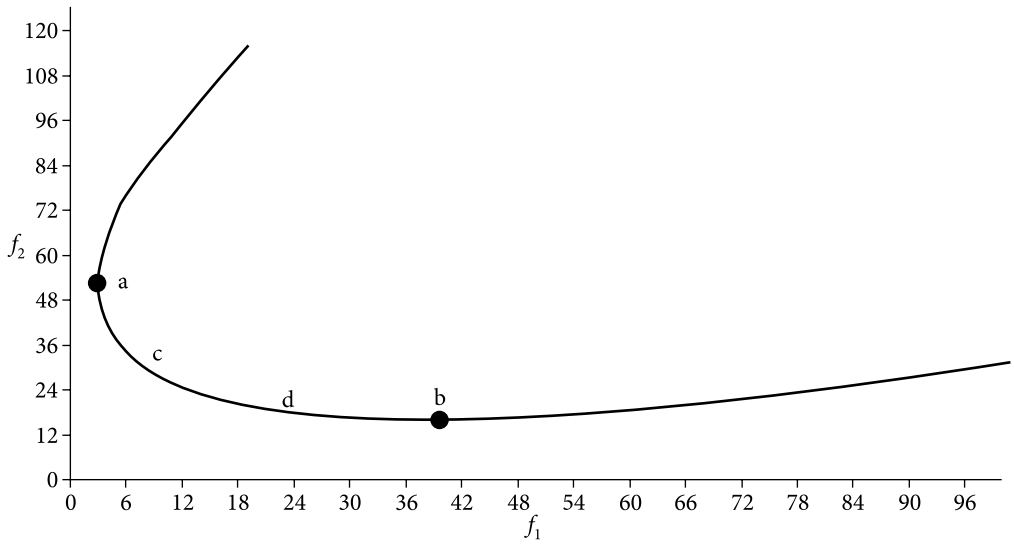


Fig. 8.3 Pareto optimal set

A very common example of MOO is simultaneous optimization of yield and selectivity of the intermediate desired product B for an isothermal batch reactor with series reaction $A \xrightarrow{k_1} B \xrightarrow{k_2} C$. We are interested to maximize both the yield and the selectivity of the desired product, *B*, simultaneously. The optimization problem for this system can be formulated as

$$\text{Max } Y_B = \frac{C_B}{C_A} \tag{8.57}$$

$$\text{Max } S_B = \frac{C_B}{C_B + C_C} \tag{8.58}$$

Maximization of the yield is required since it leads to higher amounts of *B*. Whereas the maximization of the selectivity is desired to reduce the downstream separation costs. For any feed concentration, the yield and selectivity can be formulated as a function of time (t) and temperature, by integrating the mass balance equations for this system (use similar method of section 2.4.1 for batch reactor). Figure 8.4 shows the graph between yield and selectivity.

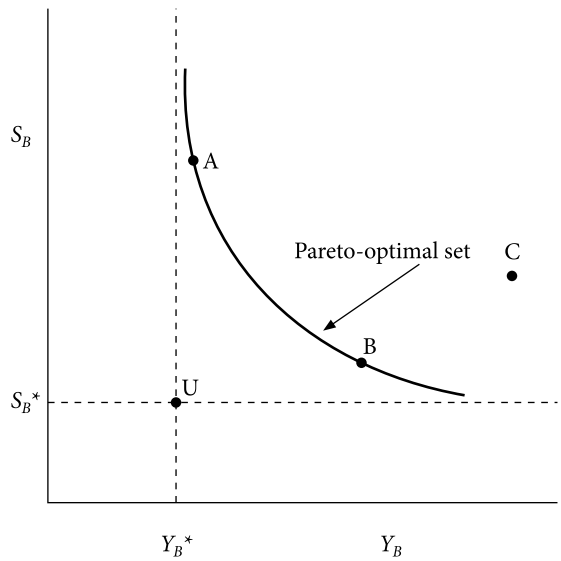


Fig. 8.4 Utopia point

Utopia point

Utopia points are points that optimize (minimize or maximize) all the objective functions of the multi-objective problem at once. This is given by the intersection of the minima of all independent objectives; this is also known as ideal point. The utopian point F^* is defined by $F^* = \left\{F_1^*, F_2^*, \ldots, F_k^*\right\}$. In Fig. 8.4, point U represents the utopia point for a bi-objective optimization problem. This point is the intersection of lines $Y_B = Y_B^*$ and $S_B = S_B^*$.

The utopia point is unachievable because the objectives are contradictory. Though, it can be utilized as a reference point for further calculation. For instance, we can find the closest point of utopia point on the Pareto front (compromise solution). Mathematically compromise solution can be represented as

$$\min_{Y_B, S_B} \left\| \Phi\left(Y_B, S_B\right) - U \right\| \tag{8.59}$$

8.2.1 Basic theory of multi-objective optimization

Although, there are multiple Pareto optimal solutions, only one solution has to be selected for practical implementation. Therefore, the multi-objective optimization process can be distinguished into two tasks: i) find a set of Pareto optimal solutions, and ii) select the most preferred solution from this set. The latter task requires a "Decision Maker" to choose the best solution in a particular instance of MOO problem as all Pareto optimal solutions are mathematically equivalent. A decision maker is one who can express preference information based on the requirement about the mathematically equivalent Pareto optimal solutions.

There are two main approaches for solving multi-objective optimization problems. The first one is Multi-Criteria Decision Making (MCDM) approach that can be characterized by the use of mathematical programming techniques and a decision making method in an intertwined manner. The decision maker plays a vital role in providing information to build a preference model. Evolutionary Optimization is a useful approach to solve MOO problems. Since evolutionary algorithms are population-based approach, they usually find an approximation of the Pareto front in a single run.

In this section, we will discuss various methods like Lexicographic Method, Linear Weighted Sum Method, Evolutionary Multi-objective Optimization, and Utopia-Tracking approach.

8.2.1.1 Lexicographic method

In this method, the objectives are ranked in order of importance (from best to worst) by the decision maker. The optimal solution X^* is found by minimizing the objective functions successively. The optimal value $F_i^*\left(i=1,2,\ldots,k\right)$ is found by minimizing the objective functions sequentially, starting with the most important objective function and proceeding according to the order of importance. Additionally, the optimal value (F_i^*) found of objective $F_i(X)$ is added as a constraint for optimization of $F_{i+1}(X)$. In this way, we move towards the optimal value of the most important objectives.

The subscripts of the objectives signify the objective function number as well as the priority of the objective. Therefore, $F_1(X)$ and $F_k(X)$ represent the most and least important objective functions, respectively. Therefore, the first problem is developed as

$$\text{Minimize } F_1\left(X\right) \tag{8.60a}$$

$$\text{subject to } g_j\left(X\right)\le 0,\ \ j=1,2,\ldots,m \tag{8.60b}$$

The solution of this problem is X_1^* and $F_1^* = F_1\left(X_1^*\right)$. After that the second problem is constructed as

$$\text{Minimize } F_2\left(X\right) \tag{8.61a}$$

$$\text{subject to } g_j\left(X\right)\le 0,\ \ j=1,2,\ldots,m \tag{8.61b}$$

$$F_1\left(X\right)=F_1^* \tag{8.61c}$$

the solution of this problem is found as X_2^* and $F_2^* = F_2\left(X_2^*\right)$.

This process repeats k times, until all the k objectives have been considered. The ith problem can be written as

$$\text{Minimize } F_i(X) \tag{8.62a}$$

$$\text{subject to } g_j(X) \le 0,\ j = 1,2,\ldots,m \tag{8.62b}$$

$$F_l(X) = F_l^*,\ l = 1,2,\ldots,i-1 \tag{8.62c}$$

and the solution of this ith problem is X_i^* and $F_i^* = F_i\left(X_i^*\right)$. Finally, the solution obtained at the end (X_k^*) is considered as the desired solution X^* of the original multi-objective optimization problem.

It is proved that the optimal solution obtained by the lexicographic problem is Pareto optimal. For this reason, the lexicographic method is usually adopted as an additional optimization approach in methods that can only guarantee weak optimality by themselves [Ehrgott, (2005); MIETTIEN, (1999)].

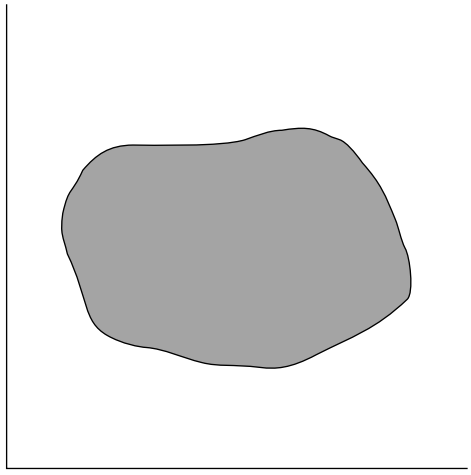
8.2.1.2 Linear weighted sum method

Multi-objective optimization can be modeled and solved by transforming it into a single objective problem using different methods, namely restriction method, ideal point method, and linear weighted sum method. These methods largely depend on the values assigned to the weighted factors or the penalties used, which are done quite arbitrarily. Another disadvantage of the above methods is that these algorithms obtain only one optimal solution at a time and may miss some useful information [Li *et al.* (2003)]. The weighted sum method is used extensively for solving MOO problems. This method converts multiple objective functions into an aggregated objective. This is done by multiplying each objective function with a weighting factor and summing up all weighted objective functions. Problem stated in Eq. (8.56b) can be written as [Kim, (2004)]

$$F_{WS} = \sum_{i=1}^{k} w_i F_i = w_1 F_1 + w_2 F_2 + \cdots + w_k F_k \tag{8.63}$$

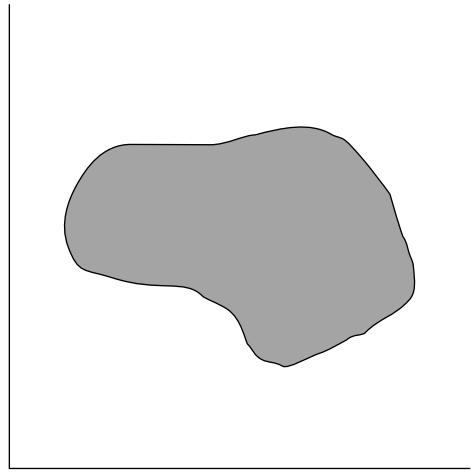
where w_i = (i = 1,2,…k) represents the weighting factor for the ith objective function (potentially also dividing each objective by a scaling factor $w_i = \alpha_i / sf_i$). The weighted sum is said to be a convex combination of objectives when $\sum_{i=1}^{k} w_i = 1$ and $0 \le w_i \le 1$. Each single objective optimization provides one particular optimal solution point on the Pareto front. The weighted sum method then modifies weights systemically, and each different single objective optimization determines a different optimal solution. The solutions obtained estimate the Pareto front [Kim, (2004)].

The major drawbacks of the weighted sum method are (1) often the distribution of the optimal solution is not uniform, and (2) this method is not suitable for non-convex regions (Figs 8.5(a), 8.5(b)). Adaptive weighted sum method has been developed to optimize non-convex objective space.



(a)

Fig. 8.5(a) Convex objective space



(b)

Fig. 8.5(b) Non-convex objective space

Finding the proper weights for Eq. (8.63) is a crucial job for the decision makers. Setting proper weights is just one way of expressing preferences, and this approach is applicable to many different methods. Misinterpretation of the practical and theoretical significance of the weights can make the method of intuitively choosing non-arbitrary weights an ineffective task [Marler and Arora, (2004)]. Therefore, understanding how the weights are affecting the solution to the weighted sum method has implications about other approaches that involve similar method parameters [Marler and Arora, (2010)]. Many researchers have developed methodical approaches for deciding weights.

In ranking methods, the objective functions are arranged according to their importance. The least important objective be given a weight of 1 and integer weights with subsequent increments are given to objective functions that are more important. The similar way is followed in categorization methods, where various objectives are segregated in broad categories for instance highly important, moderately important, and less important. The decision maker allocates independent values of relative importance to each objective function in the ranking methods [Marler and Arora, (2004)].

A study shows that values of weights can be decided using a preference function. A preference function is an abstract function, which is made by the points in the criterion space. This function perfectly incorporates the preferences, which is in the mind of the decision-maker. The majority of the MOO methods that minimizes a single combined objective function, which try to approximate the preference function with some mathematical formulation, called a utility function. The gradients of P[$F(X)$], the preference function and $U = \sum_{i=1}^{k} w_i F_i(X)$, the utility function are given in Eqs (8.64) and (8.65) respectively:

$$\nabla_X P\left[F(X)\right] = \sum_{i=1}^{k} \frac{\partial P}{\partial F_i} \nabla_X F_i(X) \tag{8.64}$$

$$\nabla_X U = \sum_{i=1}^{k} w_i \nabla_X F_i(X) \tag{8.65}$$

Each part of the gradient $\nabla_X P$ qualitatively shows how the satisfaction of decision-makers changes with the change in design point consequently change in function values. Comparing Eqs (8.64) and (8.65) shows that when the weights are chosen correctly, the utility function and preference function can have gradients that are parallel to each other. This is very significant as the purpose of any utility function is approximating the preference function, a more accurate representation of one's preferences can result by imposing similarities between these two functions such as parallel gradients. When the utility function and the preference function are ideally same, then definitely the gradients of these two functions should be the same as well.

Equation (8.64) and (8.65) show that w_i corresponds to $\partial P/\partial F_i$ value. The value of $\partial P/\partial F_i$ is the approximate change in the preference function value that results from a change in the objective function value for F_i. Thus, $\partial P/\partial F_i$ gives us a mathematical definition for the importance of F_i. It will be better if we consider the significance of an objective function or change in preference function value in relative terms. Consequently, the value of a weight is significant relative to the values of other weights; the independent absolute magnitude of a weight is not relevant in terms of preferences [Marler and Arora, (2010)].

8.2.1.3 Evolutionary multi-objective optimization

In some cases, computational cost for generating the Pareto set is high and often it is not feasible; the difficulty of the underlying application makes the exact methods inapplicable. Therefore, a number of stochastic optimization techniques like simulated annealing, tabu search, ant colony optimization, etc. have been used to generate the Pareto set. The term evolutionary algorithm (EA) is used for a class of stochastic optimization methods that replicate the process of natural evolution. These methods have been used successfully for solving MOO problem [Deb (2001)]. The EAs are able to find a complete set of Pareto optimal solutions in a single run as these algorithms simultaneously deal with a set of possible solutions (population). Whereas the conventional mathematical programming techniques, need a series of separate runs. Even though the fundamental mechanisms are simple, the EAs have established themselves as a general, robust and powerful search mechanism [Bäck *et al.* (1997)] Particularly they have some characteristics that are required for problems involving i) objectives with multiple conflicts, and ii) obstinately large and very complex search spaces. The main advantage of EAs are, they are useful when we have inadequate knowledge about the problem being solved, easy to implement, robust, could be employed in a parallel environment, and are less susceptible to the continuity or shape of the Pareto front.

The EA is initiated with a population of solution candidates. Then to generate new solutions, reproduction process is used. The reproduction process enables to combine the existing solutions. Finally, natural selection decides which individuals of the current population join in the new population. The flowchart (Fig. 8.6) and algorithm can be written as follows:

Step 1 Initialize the random population, and additional parameters

Step 2 Calculate the fitness of each individual in the population

Step 3 Choose best-ranking individuals for reproducing

Step 4 Generate a new offspring through crossover operator

Step 5 Generate a new offspring through mutation operator

Step 6 Calculate the fitness of the individual offspring

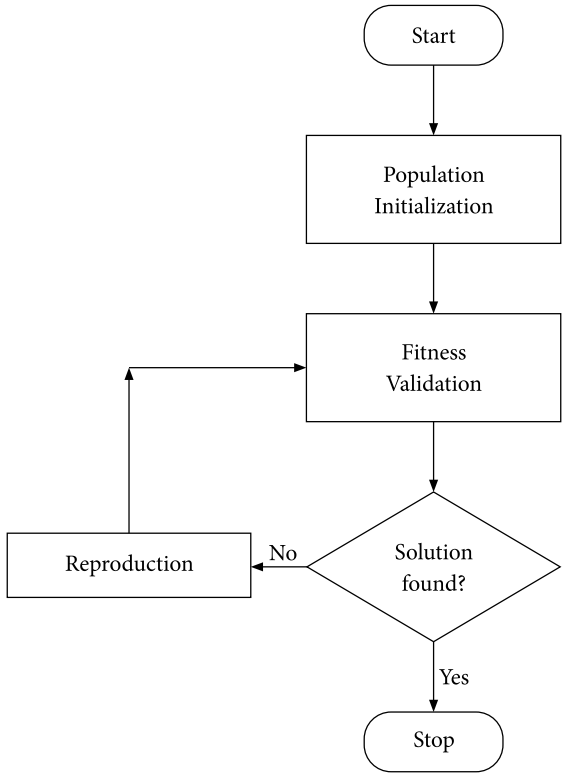


Fig. 8.6 Flowchart for evolutionary algorithm

Step 7 Replace worst ranked part of population with offspring

Step 8 Repeat, until terminating criteria matched

Details of different algorithm like genetic algorithm, simulated annealing, differential evolution have been elucidated in chapter 9.

8.2.1.4 Utopia-tracking approach

The main technical challenge in tackling the multiple objectives is that the construction of Pareto front is computationally expensive, mainly with multiple dimensions. Moreover, expert knowledge is still required to achieve a preferred solution even when such a front is built. Conventional approaches such as weighting and expert systems are limited as system conditions and priorities change under various operating modes [Zavala and Flores-Tlacuahuac, (2012)].

The key idea of the Utopia-Tracking approach is to minimize the distance between the cost function and the utopia point. This problem can be formulated as

$$\min \left\| F - U \right\|_p = \left(\sum_{i=1,2} \left| f_1 - f_1^L \right|^p + \left| f_2 - f_2^L \right|^p \right)^{1/p} \tag{8.66}$$

Here, $\|\bullet\|_p$ is the p-norm. This method is illustrated in Fig. 8.7. Where F (compromise solution) is any point on the Pareto front, U is the utopia point.

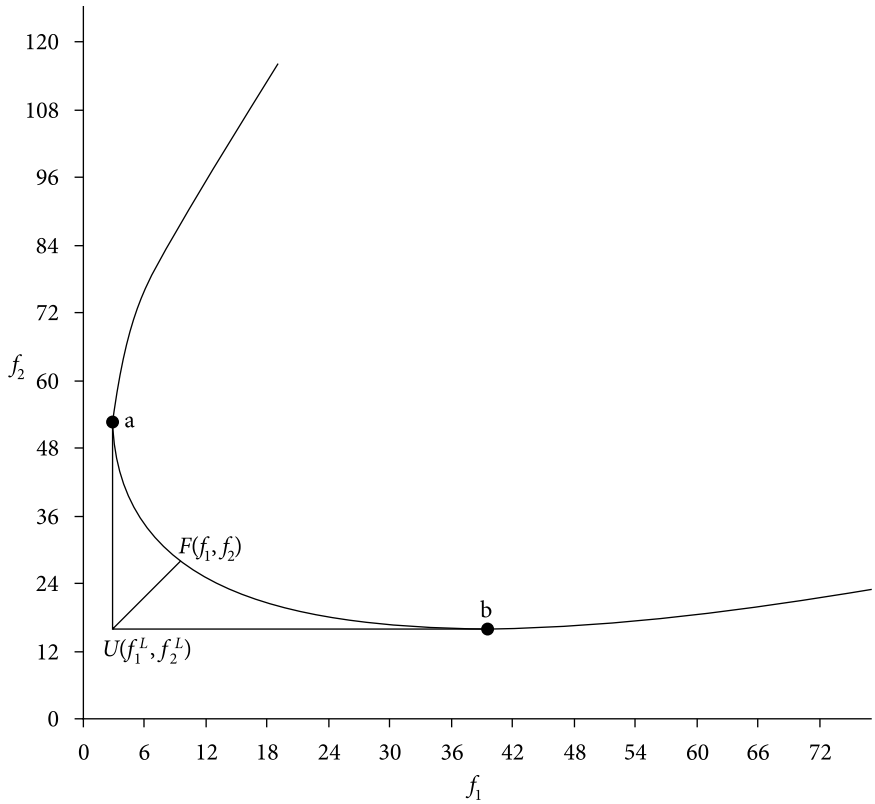


Fig. 8.7 Utopia-tracking approach

Traditional weighting method minimizes the function $w_1f_1 + w_2f_2$, where the weights are usually fixed. The weights have to be adapted to stay near to the utopia point as conditions change. The utopia-tracking is a method in which the weights of the compromise solution is determined automatically which is the closest point to the utopia point at each point in time [V.M. Zavala, (2013)]. This method has been applied successfully for optimizing control scheme like model predictive control (MPC) [Zavala and Flores–Tlacuahuac (2012)].

8.2.2 Multi-objective optimization applications in chemical engineering

Multi-objective optimization is very useful during the design of process equipments. In this chapter, multi-objective optimization of the fed batch reactor, PID controller tuning, and optimization of catalytic reforming process have been discussed.

8.2.2.1 Fed-batch bioreactor

To improve real-time decision making efficient approaches are required to determine the Pareto set in a rapid and precise way.

We consider the problem of finding the optimal feeding rate and batch duration for a fed-batch lysine fermentation process as described by Logist *et al.* [Logist *et al.* (2009)]. The state variables and the control variable of the process are given as follows.

x_1 : biomass [*g*],

x_2 : substrate [*g*],

x_3 : product, lysine [*g*],

x_4 : fermenter volume [*l*],

u : volume rate of the feed stream [*l*/*h*],

Two competing objective functionals are to be maximized, in search for Pareto-optimal feeding strategy, $u(t)$, and batch duration, t_f[*h*] .

i. The ratio of the product formed and the process duration, i.e., the productivity:

$$-\varphi_1\left(x\left(t_f\right),t_f\right)=\frac{x_3\left(t_f\right)}{t_f} \tag{8.67}$$

ii. The ratio of the product formed and the mass of the substrate, i.e., the yield:

$$-\varphi_2\left(x\left(t_f\right),t_f\right)=\frac{x_3\left(t_f\right)}{2.8\left(x_4\left(t_f\right)-5\right)} \tag{8.68}$$

Note that clearly t_f is free, but, it is subject to the constraint

$$20\le t_f\le 40 \tag{8.69}$$

Constraints are also imposed on the fermenter volume, the feed rate and the amount of substrate to be added, respectively, as follows.

$$5\le x_4\left(t\right)\le 20 \tag{8.70a}$$

$$0\le u\left(t\right)\le 2 \tag{8.70b}$$

$$20\le 2.8\left(x_4\left(t_f\right)-5\right)\le 42 \tag{8.70c}$$

The objective functions, constraints, and the process equations as given in Logist *et al.* [Logist *et al.* (2009)] yield the following bi-objective optimal control problem.

$$\min \left(-\frac{x_3\left(t_f\right)}{t_f}, \frac{x_3\left(t_f\right)}{2.8\left(x_4\left(t_f\right)-5\right)} \right) \tag{8.71a}$$

subject to

$$\dot{x}_1 = \left(0.125\frac{x_2}{x_4}\right)x_1, \quad x_1(0) = 0.1 \tag{8.71b}$$

$$\dot{x}_2 = -\left(\frac{0.125}{0.135}\frac{x_2}{x_4}\right)x_1 + 2.8u, \quad x_2(0) = 14 \tag{8.71c}$$

$$\dot{x}_3 = \left[-384\left(0.125\frac{x_2}{x_4}\right)^2 + 134\left(0.125\frac{x_2}{x_4}\right)\right]x_1, \quad x_3(0) = 0 \tag{8.71d}$$

$$\dot{x}_4 = u, \quad x_4(0) = 5 \tag{8.71e}$$

$$0 \le u(t) \le 2,\ 5 \le x_4(t) \le 20,\ 20 \le t_f \le 40,\ 20 \le 2.8\left(x_4\left(t_f\right)-5\right) \le 42 \tag{8.71f}$$

8.2.2.2 MOO for PID controller tuning

PID controller tuning can be formulated as a multi-objective optimization problem. During controller tuning we need to find the parameters (e.g., k_P, τ_I, τ_D) of any PID controller. For this purpose, we need to minimize the error functions like ISE, IAE, and ITAE. These error functions are represented as

Integral Square Error (ISE): $$\text{ISE} = \int_0^\infty \varepsilon^2(t)dt \tag{8.72}$$

Integral Absolute Error (IAE): $$\text{IAE} = \int_0^\infty |\varepsilon(t)|dt \tag{8.73}$$

Integral of the Time-weighted Absolute Error (ITAE): $$\text{ITAE} = \int_0^\infty t|\varepsilon(t)|dt \tag{8.74}$$

The optimization scheme has been shown in Fig. 8.8

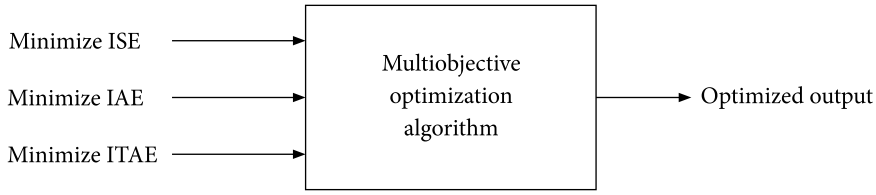


Fig. 8.8 PID controller as MOO problem

Mathematically, it can be written as

$$\text{Min. ISE} = \int_0^\infty \varepsilon^2(t)dt \tag{8.75a}$$

$$\text{Min. IAE} = \int_0^\infty |\varepsilon(t)|dt \tag{8.75b}$$

$$\text{Min. ITAE} = \int_0^\infty t|\varepsilon(t)|dt \tag{8.75c}$$

8.2.2.3 Industrial naphtha catalytic reforming process

The variables that affect the catalytic reforming process are the volume flow of naphtha charge to the volume of the catalyst (liquid hourly space velocity, LHSV), the latent aromatics content of naphtha charge (LA), the inlet temperatures of four reactor (T_1, T_2, T_3, T_4), the reaction pressure (p_r), the mole flow of hydrogen in the recycle gas to the mole flow of naphtha charge (hydrogen to oil molar ratio, n_{H_2}/n_{HC}), the product separator temperature (T_s), etc. Among these nine process variables chosen using mechanism analysis, the sensitivity analysis of each variable is performed using the process. It is shown that the appropriate set point value of one variable for maximizing the aromatics yield may not be suitable for minimizing the yield of heavy aromatics. So, the suitable trade-off solutions for the two optimal objectives should be considered.

For the continuous catalytic reforming process in this study, the unit is in full load operation and the value of LHSV cannot be further increased. Similarly, the quality of naphtha feedstock (e.g., LA) cannot be changed artificially for most domestic petroleum-refining enterprises. The product separator temperature T_s is not independent of other variables. Moreover, for further lowering of the temperature, coolers need to be included in the system, which in turn increase the operation costs. Hence, the remaining process variables are chosen as the decision variables for optimization of this process. These are the four reactor inlet temperatures (T_1, T_2, T_3, T_4), the reaction pressure (p_r), and the hydrogen-to-oil molar ratio (n_{H_2}/n_{HC}).

Thus, the two independent objectives, namely, the maximization of the aromatics yield (AY) and the minimization of the yield of heavy aromatics (HAY) are formulated mathematically as follows:

$$\text{maximize AY}\left(T_1, T_2, T_3, T_4, p_r, n_{H_2}/n_{HC}\right) \tag{8.76a}$$

$$\text{minimize } \text{HAY}\left(T_1, T_2, T_3, T_4, p_r, n_{H_2}/n_{HC}\right) \tag{8.76b}$$

$$\text{subject to } 520 \le T_1, T_2, T_3, T_4 \le 530 \tag{8.76c}$$

$$0.8 \le p_r \le 0.9 \tag{8.76d}$$

$$3.0 \le n_{H_2}/n_{HC} \le 4.0 \tag{8.76e}$$

$$65 \le \text{AY} \le 68 \tag{8.76f}$$

$$18 \le \text{HAY} \le 23 \tag{8.76g}$$

Weifeng *et al.* [Weifeng *et al.* (2007)] used neighborhood and archived genetic algorithm (NAGA) to this MOO problem.

8.3 Optimization in Control Engineering

Application of optimization techniques in the field of control engineering is very popular in recent times. Tuning of PID controller, application of MPC, and designing of optimal control systems need sufficient knowledge of optimization techniques. Some applications are discussed in the following section.

8.3.1 Real time optimization

The online calculation of optimal set points is known as Real Time Optimization (RTO). RTO deals with techniques of maximizing an economic objective related to the operation of a continuous process while satisfying other operating constraints [De Souza *et al.* (2010)]. Most RTO systems require non-linear steady state models of the process system combined with data reconciliation. This data reconciliation system updates the major parameters such as feed compositions and efficiencies using some parameter estimation techniques [Marlin and Hrymak (1996)]. The RTO system optimizes the process operating conditions (i.e., temperature, pH) and updates the set points to local MPCs that are developed based on linear dynamic models. If the set of active constraints with significant economic importance changes frequently, then a steady state RTO may not be sufficient. For these processes, it is more appropriate to use dynamic optimization with a nonlinear model developed by using dynamic RTO (DRTO) or nonlinear MPC with economic objective [Tosukhowong *et al.* (2004]. Even though for practical application, steady state RTO is more appropriate than the DRTO.

As RTO relies on a static model, the plant should reach at sufficiently steady state to update the plant model properly. Typical steady-state detection methods include statistical parameters like mean, variance, or slope for selected signals over a moving window [Shrikant and Saraf (2004)].

Implementation of RTO system in process industries is a big challenge. Despite of many advantages, some companies feel that RTO is not viable to them. However, some companies are convinced of the profit from RTO systems and continue to invest in research and practical applications [Darby *et al.* (2011)].

Figure 8.9 shows the Plant decision hierarchy and approximate time level for different steps.

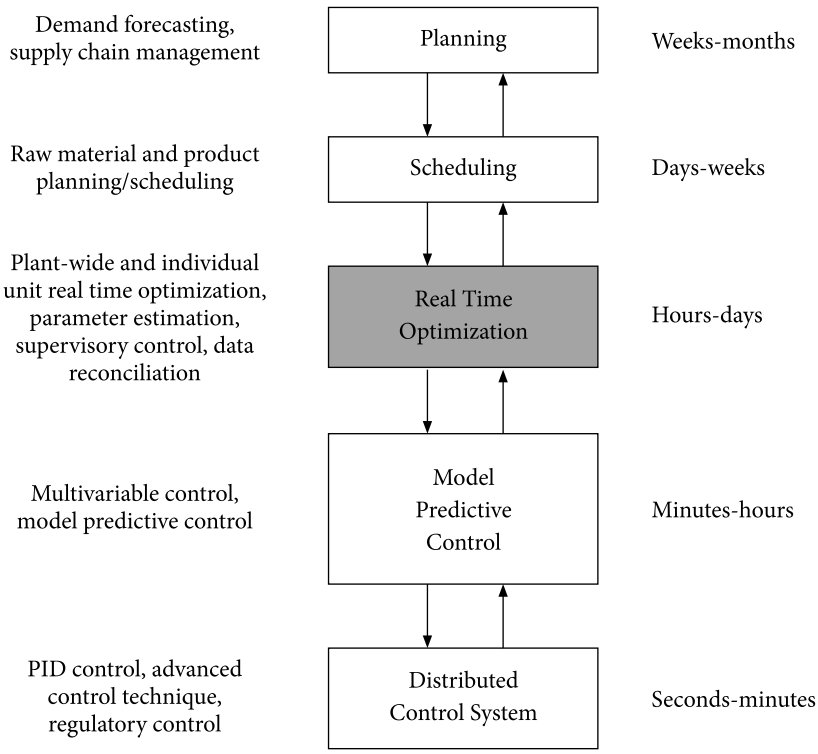


Fig. 8.9 Plant decision hierarchy

However, the execution rate of RTO depends on two factors; the frequency of unmeasured disturbances and the time necessary for an MPC to move the process to a new steady state. There is no clear-cut distinction of time scales, i.e., open-loop response times are similar because both consider the same variables (e.g., compositions). The MPC first discard unmeasured disturbances in those variables before a new steady state can be attained [Darby *et al.* (2011)].

Modern industrial practice validates input data primarily by using bound checks, rather than by statistical data reconciliation/gross error detection procedures. When this validity check fails, alternate measurements or estimates may be replaced automatically if it is accessible. For significant variables, if appropriate alternative value is not available for replacement, the implementation of the optimizer is aborted until suitable data is available. To minimize the deviation between measured variables and model variables, the RTO model is updated (calibrated) at each cycle. Model parameters such as heat transfer coefficients, distillation efficiencies, and unmeasured flows are usually adjusted to achieve this objective. When the model is updated properly, then optimization is executed. Usually, the objective function is developed based on profit. After successful convergence

of the optimization algorithm, the optimal decisions are passed to MPC. Most of the industrial MPC applications work based on linear models, which are build empirically from a dedicated plant test [Darby *et al*. 2009]. Linear models (or linearized transform) have shown to be satisfactory for numerous control problems encountered in the chemical process industries [Gattu *et al*. 2003].

There are different techniques to accommodate RTO set points for controlled and manipulated variables in MPC. To penalize deviations from RTO set points, an additional term is included in the MPC objective function. Another formulation is done by using a separate target selection layer to determine the best, feasible set points for the MPC [Muske and Rawlings, (1993); Ying and Joseph (1999)]. We are considering the target selection is a linear programming (LP) optimizer, which is executing at a same frequency of the MPC that is common in the industrial application. The main advantage of this two-stage implementation over the single stage formulation is that it can better respond to disturbances and maintain feasibility between RTO executions [Ying and Joseph (1999)]. The following form can represent the typical LP formulation:

$$\min_{y^{TG},u^{TG}} c_y^T y^{TG} + c_u^T u^{TG} \tag{8.77a}$$

subject to

$$y^{TG} = G_{ss} u^{TG} + b \tag{8.77b}$$

$$y^{\min} \le y^{TG} \le y^{\max} \tag{8.77c}$$

$$u^{\min} \le u^{TG} \le u^{\max} \tag{8.77d}$$

where c_y and c_u represent the costs associated with the outputs and inputs (controller manipulated variables), respectively; y^{TG} and u^{TG} are the optimal targets for the outputs and inputs, respectively; superscripts min and max give the limits of constraint for the inputs and outputs; G_{ss} re[resents the steady-state gain matrix obtained from the dynamic model of MPC; and b is the steady-state model bias term that is updated based on present output measurements and predictions from the dynamic model.

8.3.2 Optimal control of a batch reactor

In this section we will discuss the optimal control of a batch reactor used for bio biodiesel production. This optimization scheme was developed by Benavides and Diwekar [Benavides and Diwekar, 2012].

Batch reactor model

Biodiesel was produced by the transesterification of triglycerides and methanol in the presence of an alkaline catalyst for example sodium hydroxide. This reversible reaction composed of 3 steps, where triglycerides (TG) are converted to diglycerides (DG), diglycerides to monoglycerides

(MG), and lastly monoglycerides to glycerol (GL). An ester (R_iCOOCH_3) is produced in each step; resulting in 3 molecules of ester from one molecule of triglycerides. Steps of the reaction are given by Eqs (8.78a)–(8.78c) where k_1 to k_6 are the rate constants; and the overall reaction is represented by Eq. (8.79). All reactions are taken place at atmospheric pressure [Noureddini and Zhu (1997)].

$$TG + CH_3OH \xleftrightarrow{k_1,k_2} DG + R_1COOCH_3 \tag{8.78a}$$

$$DG + CH_3OH \xleftrightarrow{k_3,k_4} MG + R_2COOCH_3 \tag{8.78b}$$

$$MG + CH_3OH \xleftrightarrow{k_5,k_6} GL + R_3COOCH_3 \tag{8.78c}$$

the overall reaction is given by the following equation

$$TG + 3CH_3OH \longleftrightarrow 3RCOOCH_3 + GL \tag{8.79}$$

The mathematical model of this batch reactor is represented by the following Ordinary Differential Equations (ODEs) (8.80)–(8.85) obtained from the mass balance of the batch reactor [Noureddini and Zhu (1997)].

$$F_1 = \frac{dC_{TG}}{dt} = k_1 C_{TG} C_A + k_2 C_{DG} C_E \tag{8.80}$$

$$F_2 = \frac{dC_{DG}}{dt} = k_1 C_{TG} C_A - k_2 C_{DG} C_E - k_3 C_{DG} C_A + k_4 C_{MG} C_E \tag{8.81}$$

$$F_3 = \frac{dC_{MG}}{dt} = k_3 C_{DG} C_A - k_4 C_{MG} C_E - k_5 C_{MG} C_A + k_6 C_{GL} C_E \tag{8.82}$$

$$F_4 = \frac{dC_E}{dt} = k_1 C_{TG} C_A - k_2 C_{DG} C_E + k_3 C_{DG} C_A - k_4 C_{MG} C_E + k_5 C_{MG} C_A - k_6 C_{GL} C_E \tag{8.83}$$

$$F_5 = \frac{dC_A}{dt} = -\frac{dC_E}{dt} \tag{8.84}$$

$$F_6 = \frac{dC_{GL}}{dt} = k_5 C_{MG} C_A - k_6 C_{GL} C_E \tag{8.85}$$

where C_{TG}, C_{DG}, C_{MG}, C_E, C_A, and C_{GL} are concentrations of triglycerides, diglycerides, monoglycerides, methyl ester, methanol, and glycerol respectively.

Optimal control problem

Optimal control provides us useful information for designing and controlling the reaction process. It has numerous applications in the academic and industrial field. Usually, a time dependent profiles of the control variable are required to optimize a particular performance index [Rico-Ramirez and Diwekar (2004)]. For example, one of the most common cases of an optimal control problem is that the best temperature profile is found so that the performance index, conversion or yield, is optimized at a particular final time. Some other methods can be employed to solve these problems are the dynamic programming, maximum principle, and calculus of variations [Diwekar (2003)].

These methods depend on suitable mathematical representations, which cannot be solved directly by simple mathematical models. Methods with second order differential equations or partial differential equations are difficult to solve so the maximum principle that uses ODEs, can be employed in most of cases [Diwekar (1995)]. At each time step, the maximum principle method adds adjoint variables and corresponding adjoint differential equations, and a Hamiltonian that requires to be optimized to get the control variable. In this section, an optimal control problem is formulated for the production of biodiesel in a batch reactor. Generally two different types of optimization problems are considered: the maximization of concentration and minimization of time. Though, only the maximum concentration is discussed in this section. We need to maximize the biodiesel concentration. For this optimal control problem, the best temperature profile in a given reaction time (100 min) is found to maximize the concentration of methyl ester (biodiesel). The maximum principle formulation is used for solving this problem. The objective function is reconstructed in this method. It is represented as a linear function in terms of final values of state variables (C_i) and the constant values (A_i), therefore, the objective function for this problem is given below:

$$\text{Maximize } J = \sum_{i=1}^{n} A_i C_i\left(t_f\right) = \overline{A}_i^T . \overline{C}_i\left(t_f\right) = C_E\left(t_f\right) \tag{8.86a}$$

subject to state Eqs (8.80) to (8.85) given in the generalized form below

$$\frac{dC_i}{dt} = f\left(C_i, T\right) \tag{8.86b}$$

for

$$C_i = \left[C_{TG}\left(t_0\right), C_{DG}\left(t_0\right), C_{MG}\left(t_0\right), C_E\left(t_0\right), C_A\left(t_0\right), C_{GL}\left(t_0\right)\right] \tag{8.86c}$$

where C_i denotes the state variable that corresponds to the concentration of each component: C_{TG}, C_{DG}, C_{MG}, C_E, C_A, and C_{GL}. Temperature T is a control variable. A_i represents the constants values for the linear representation of the maximum principle. Initial conditions:

Initial time (t_0) = 0 min.

$$C_i\left(t_0\right) = \left[0.3226; 0; 0; 0; 1.9356; 0\right] \text{ mol/l}$$

$$A = [0;0;0;1;0;0]$$

The maximum principle employs the addition of n adjoint variables (one adjoint variable per state variable), n adjoint equations, and a Hamiltonian that satisfies the following relations

$$H\left(\overline{z}_t, \overline{C}_t, T\right) = \overline{z}_t^T, F\left(\overline{C}_t, T_t\right) = \sum_{i=1}^{n} z_i F_i\left(\overline{C}_t, T_t\right) \tag{8.87}$$

$$\frac{dz_i}{dt} = \sum_{j=1}^{n} z_j \left(\frac{\partial F_i}{\partial C_i}\right) \tag{8.88}$$

where n is the number of components (six) and F_i is the right-hand side of the differential Eqs (8.90a) to (8.90f). For this problem Hamiltonian can be written as

$$H = z_1 F_1 + z_2 F_2 + z_3 F_3 + z_4 F_4 + z_5 F_5 + z_6 F_6 \tag{8.89}$$

The adjoint equations can be compute as

$$\frac{dz_1}{dt} = -z_1\left(-k_1 C_A\right) - z_2\left(k_1 C_A\right) - z_4\left(k_1 C_A\right) - z_5\left(-k_1 C_A\right) \tag{8.90a}$$

$$\begin{aligned}\frac{dz_2}{dt} = &-z_1\left(k_2 C_A\right) - z_2\left(-k_2 C_E - k_3 C_A\right) - z_3\left(k_3 C_A\right) \\ &- z_4\left(-k_2 C_E + k_3 C_A\right) - z_5\left(k_2 C_E - k_3 C_A\right)\end{aligned} \tag{8.90b}$$

$$\begin{aligned}\frac{dz_3}{dt} = &-z_2\left(k_4 C_E\right) - z_3\left(-k_4 C_E - k_5 C_A\right) - z_4\left(-k_4 C_E + k_5 C_A\right) \\ &- z_5\left(k_4 C_E - k_5 C_A\right) - z_6\left(k_5 C_A\right)\end{aligned} \tag{8.90c}$$

$$\begin{aligned}\frac{dz_4}{dt} = &-z_1\left(k_2 C_{DG}\right) - z_2\left(-k_2 C_{DG} + k_4 C_{MG}\right) - z_3\left(-k_4 C_{MG} + k_6 C_{GL}\right) \\ &- z_4\left(-k_2 C_{DG} - k_4 C_{MG} - k_6 C_{GL}\right) - z_5\left(k_2 C_{DG} + k_4 C_{MG} + k_6 C_{GL}\right) + z_6\left(k_6 C_{GL}\right)\end{aligned} \tag{8.90d}$$

$$\begin{aligned}\frac{dz_5}{dt} = &\, z_1\left(k_1 C_{TG}\right) - z_2\left(k_1 C_{TG} - k_3 C_{DG}\right) - z_3\left(k_3 C_{DG} - k_5 C_{MG}\right) \\ &- z_4\left(k_1 C_{TG} + k_3 C_{DG} + k_5 C_{MG}\right) - z_5\left(-k_1 C_{TG} - k_3 C_{DG} - k_5 C_{MG}\right) - z_6\left(k_5 C_{MG}\right)\end{aligned} \tag{8.90e}$$

$$\frac{dz_6}{dt} = -z_3\left(k_6 C_E\right) - z_4\left(k_6 C_E\right) - z_5\left(k_6 C_E\right) + z_6\left(k_6 C_E\right) \tag{8.90f}$$

For the adjoint variable, the boundary conditions are $z_i(t_f)$ = [0;0;01;0;0], which stands for the constant values of vector A. These equations can be solved easily using backward integration along with Runge–Kutta–Fehlberg (RKF) method. At last, the optimal decision vector $T(t)$ can be achieved by finding the extremum of the Hamiltonian at each time step, in other words, applying the optimality condition:

$$\left.\frac{dH}{dT}\right|_t = 0 \tag{8.91}$$

As it was stated earlier, the optimal control problem can also be described as the minimization of time problem. The objective is to minimize the batch time for a given final concentration. The objective function can be described as given in Eq. (8.92a) and the differential equations (Eqs (8.80)–(8.85)) can be converted by multiplying with dt/dC_E, as given away in Eq. (8.92b):

$$\text{Minimize } J = \int_{C_E(t_0)}^{C_E(t_f)} \frac{dt}{dC_E} = t_f \tag{8.92a}$$

$$\text{subject to } \frac{dC_i}{dC_E} = \frac{dC_i}{dt} \Big/ \frac{dC_E}{dt} \tag{8.92b}$$

These equations can be solved using *Runge–Kutta–Fehlberg method* (RK45).

8.3.3 Optimal regulatory control system

Consider a process described by a kth order linear differential equation,

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + \cdots + a_k y_{n-k} + b_1 m_{n-1} + \cdots b_k m_{n-k} \tag{8.93}$$

where $a_1, a_2, \ldots, a_k; b_1, \ldots, b_k$ are constant parameters with known values. Suppose that the purpose of the regulatory control system is to keep the closed-loop output as close as possible to a prescribed set point value y_{SP} in the presence of disturbance changes. The deviation can be specified by one of the following measures:

$$P_1 = \{y_n - y_{SP}\}^2 \tag{8.94}$$

$$P_2 = \frac{1}{N}\sum_{n=1}^{N}\{y_n - y_{SP}\}^2 \tag{8.95}$$

The first measure, P_1, is referred to as one-stage control; the second as N-stage control. Thus, the controller design problem can be formulated as follows:

Find a controller that minimizes P_1 or P_2 in the presence of load changes.

The control action that minimizes P_1 attempts to keep the output close to the set point by making individual control decisions at each stage. Minimum of P_2, on the other hand, relaxes the restriction above and plans the control action over a longer time horizon. Let us now solve the two design problems above.

Suppose that we are at the nth sampling instant and that we want to compute the control action m_n in such a way that y_{n+1} will be as close as possible to the desired y_{SP}. Using criterion P_1, we take

$$P_1 = \{y_n - y_{SP}\}^2 = \{a_1 y_n + a_2 y_{n-1} + \cdots a_k y_{n-k+1} + b_1 m_n + b_2 m_{n-1} + \cdots b_k m_{n-k+1} - y_{SP}\}^2 \tag{8.96}$$

The minimum of P_1 is found when $\partial P_1 / \partial m_n = 0$. Then, we have

$$2\{a_1 y_n + a_2 y_{n-1} + \cdots a_k y_{n-k+1} + b_1 m_n + b_2 m_{n-1} + \cdots b_k m_{n-k+1} - y_{SP}\}(b_1) = 0 \tag{8.97}$$

and the optimum regulatory control action at the nth instant is given by

$$m_n = \frac{1}{b_1}\{y_{SP} - a_1 y_n - a_2 y_{n-1} - \cdots - a_k y_{n-k+1} - b_2 m_{n-1} - \cdots - b_k m_{n-k+1}\} \tag{8.98}$$

The controller defined by Eq. (8.98) is physically realizable because it uses only current or past information on the manipulated variable and the controlled output.

Now let us turn our attention to the second design criterion, P_2. Consider the situation at the (N – 1) sampling instant. The outputs y_{N-1}, y_{N-2}, …, y_1 have been measured and the control problem is to determine the value of the manipulated variable m_{N-1}. Since m_{N-1} influence only the last term of P_2, we have

$$\min P_2 = \{y_N - y_{SP}\}^2 \tag{8.99}$$

Then the optimal value of m_{N-1} is given by Eq. (8.98) with $n = N - 1$.

Consider, the situation at $n = N - 2$. The output has been measured for $n = N - 2, N - 3, \ldots, 1$ and the problem is to determine the optimal value of m_{N-2}. Since m_{N-2} influences the last two terms of P_2, we have

$$\min P_2 = \{y_N - y_{SP}\}^2 + \{y_{N-1} - y_{SP}\}^2 \tag{8.100}$$

If the optimum value of m_{N-1} has been used for the last stage, the minimization problem Eq. (8.100) yields

$$\min P_2 = \{y_{N-1} - y_{SP}\}^2 \tag{8.101}$$

because $y_N - y_{SP} = 0$ for the optimum value of m_{N-1}. However, the optimum value of m_{N-2} solving the last problem is given again by Eq. (8.98) for $n = N - 2$. Therefore, we reach the following conclusion:

The optimal regulatory control action for a system described by a *k*th order difference model with constant and known parameters is given by Eq. (8.98) independently of which criterion, P_1 or P_2 is used.

8.3.4 Dynamic matrix control

Dynamic Matrix Control (DMC) was developed by Shell Oil Company in the 1970s. Afterward, DMC has been successfully used in industry for more than a decade. Several authors have reported improved control performance by using the DMC as compared to "traditional" control algorithms [Cutler and Ramaker, (1980)]. DMC belongs to the family of model predictive control (MPC) algorithms. This algorithm can be separated into two parts, a predictor and an optimizer [Lundstrom *et al.* (1995)].

The key features of any MPC are given below:

i. linear controller
ii. uses a process or plant step response model
iii. predicts the process dynamics over a time window
iv. minimizes an objective function
v. compensates for a relatively high degree of process non-linearity (good robustness)
vi. easily adapted to multivariable plants
vii. easy tuning

Dynamic matrix control is a popular technique for the control of dynamic systems, which has slow response. The most available predictive control algorithms are implemented by optimization method that minimizes a performance index. This result in the complexity to analyze the algorithms of stability of the predictive control consequently makes it difficult to implement the stability design of predictive controllers. The calculation of inverse matrix is associated in dynamic matrix control (DMC) and it holds down on-line application. The main advantage of DMC is its capability to handle control problem with constraints.

The objective of the DMC controller is to drive the output to track the set point in the least squares sense including a penalty term on the manipulated variable moves. This results in smaller computed input moves and a less aggressive output response [Qin *et al.* (2003)]. The method is strictly developed for linear systems (like any other conventional controller) and as a result, any analysis on DMC and its features must be made in a framework of linear system theory [Garcia and Morshedi (1986)].

Linear Input-Output Model

Without loss of generality, we can consider a dynamic linear system with an input I and one output O. In computer applications, only the system behavior at the sampling intervals is of interest. A discrete representation of the dynamics is utilized for this purpose. Such a representation is given in Eq. (8.102).

$$O(k+1) = \sum_{i=1}^{M} a_i \Delta I(k-i+1) + O_0 + d(k+1) \tag{8.102}$$

where k represents discrete time; O_0 is initial condition of the output; $\Delta I(k)$ denotes the change in input (or manipulated variable) at different time intervals k; $O(k)$ represents the value of the controlled variable at time k; $d(k)$ accounts for unmodelled factors that influence $O(k)$; a_i denote the coeficients of unit step response of the system; and M is the number of time intervals necessary for the system to attain steady-state.

Therefore, $a_i = a_M$, for $i \geq M$.

The term $d(k + 1)$ has been added to the input-output description to take into consideration the unmodelled effects on the measured output. This is consisting of unmeasured disturbances and/or modelling errors. Addition of this factor is vital during the formulation of DMC as shown below.

Controller Design

The aim of any controller is to find the change in the manipulated variables, $\Delta I(k)$, which give the output $O(k)$ best match a target value O_s in the face of disturbances. Considering the present time interval to be $\bar{k}$, in DMC a projection of the output $O(k)$ over P future time intervals $\left(\bar{k}+1 \text{ to } \bar{k}+P\right)$ is matched to the set-point O_s by recommending a series of future moves.

Formulation of the DMC equations Considering Eq. (8.102), the projected output for any future time $\bar{k}+l$, $l>0$ is

$$\begin{aligned} O\left(\bar{k}+l\right) &= \sum_{i=1}^{l} a_i \Delta I\left(\bar{k}+l-i\right) \\ &+ O_0 + \sum_{i=l+1}^{M} a_i \Delta I\left(\bar{k}+l-i\right) \\ &+ d\left(\bar{k}+l\right) \end{aligned} \tag{8.103}$$

For simplicity, let us define

$$O^*\left(\bar{k}+l\right) = O_0 + \sum_{i=l+1}^{M} a_i \Delta I\left(\bar{k}+l-i\right) \tag{8.104}$$

to be the contribution to $O\left(\bar{k}+l\right)$ owing to past input moves up to the present time $\bar{k}$. This term can always be calculated from the past history of moves.

This definition can be used to write Eq. (8.103) for times $\bar{k}+1$ up to $\bar{k}+P$ to generate a set of P equations for the output projections as given below:

$$\begin{bmatrix} O\left(\bar{k}+1\right) \\ \vdots \\ O\left(\bar{k}+P\right) \end{bmatrix} = \begin{bmatrix} O^*\left(\bar{k}+1\right) \\ \vdots \\ O^*\left(\bar{k}+P\right) \end{bmatrix} + A \begin{bmatrix} \Delta I\left(\bar{k}\right) \\ \vdots \\ \Delta I\left(\bar{k}+N-1\right) \end{bmatrix} + \begin{bmatrix} d\left(\bar{k}+1\right) \\ \vdots \\ d\left(\bar{k}+P\right) \end{bmatrix} \tag{8.105}$$

where

$$A = \begin{bmatrix} a_1 & 0 & 0 & 0 \\ a_2 & a_1 & 0 & 0 \\ \vdots & & \ddots & 0 \\ a_N & a_{N-1} & & a_1 \\ \vdots & & & \\ a_M & a_{M-1} & & a_{M-N+1} \\ \vdots & \vdots & & \vdots \\ a_M & a_M & \cdots & a_M \end{bmatrix}$$

This matrix is termed as the "dynamic matrix" of the system. It is clear that during the formulation of DMC, only N moves are calculated, i.e.,

$$\Delta I(k) = 0 \quad \text{for } k > \bar{k} + N \tag{8.106}$$

Setting the values of these moves equal to zero imparts significant stability properties to the resulting controller. Particularly, we can tell from our experience that selecting $P = N + M$ yields a stable controller in most of the cases.

Estimation of unmodelled effects *d*(*k*) in DMC

An estimation of the unmodelled effects $d(k)$ is required to solve the set of equations in (8.105). As future values of the "disturbance" $d(k)$ are unavailable, the best we can do is to use an estimate. From Eq. (8.102) for $k = \bar{k} - 1$, and Eq. (8.104) for $l = 0$, we get

$$O\left(\bar{k}\right) = O^*\left(\bar{k}\right) + d\left(\bar{k}\right) \tag{8.107}$$

Therefore, $d(k)$ can be estimated using the current feedback measurement $O_m\left(\bar{k}\right)$ of O in conjunction with the moves information of past input. When we do not have any additional knowledge of $d(k)$ over future intervals (as is true in most cases), it is assumed that the predicted disturbance and the present, "measured" $d\left(\bar{k}\right)$ are equal.

$$d\left(\bar{k} + l\right) = d\left(\bar{k}\right)$$

$$= O_m\left(\bar{k}\right) - O^*\left(\bar{k}\right);\ l = 1, \ldots, P \tag{8.108}$$

Solution of the DMC equations

For a known set of equations the DMC control problem is described as finding the N future input moves $\Delta I\left(\bar{k}\right) \cdots \Delta I\left(\bar{k} + N - 1\right)$ such that the sum of squared deviations between the projections $O\left(\bar{k} + l\right)$ and the target O_s are minimized. This is same as the least-squares solution of the DMC equations:

$$\begin{bmatrix} O_s - O^*\left(\bar{k}+1\right) - d\left(\bar{k}\right) \\ \vdots \\ O_s - O^*\left(\bar{k}+P\right) - d\left(\bar{k}\right) \end{bmatrix} = e\left(\bar{k}+1\right) = AX\left(\bar{k}\right) \tag{8.109}$$

where $e\left(\bar{k}+1\right)$ represents a P-dimensional vector of projected deviations from the target and

$$X\left(\bar{k}\right) = \left[\Delta I\left(\bar{k}\right) \cdots \Delta I\left(\bar{k}+N-1\right)\right]^T \tag{8.110}$$

is the vector of future moves. Such least-squares solution can be written as

$$X\left(\bar{k}\right) = \left(A^T A\right)^{-1} A^T e\left(\bar{k}+1\right) \tag{8.111}$$

In Dynamic Matrix Control, only the move estimated for the current time interval $\bar{k}$ is employed. At every sampling time k, the computation is repeated. We obtain a new feedback measurement that is used to update $e\left(\bar{k}+1\right)$. The disturbance handling characteristics of the algorithm could be impaired, if it fails to estimate a move at each sampling time.

Formulation for multivariable systems

It should be mentioned that for a multivariable system the DMC equations could be derived in the same way as for the single-input single-output (SISO) system. For a s-input and r-output system, a linear dynamic system can be represented by

$$O\left(k+1\right) = \sum_{i=1}^{M} a_i \Delta I\left(k-i+1\right) + O_0 + d\left(k+1\right) \tag{8.112}$$

where $O(k)$ denotes an r-dimensional vector of outputs, a_i is an $r \times s$ matrix of unit step response coefficients for the ith time interval, $\Delta I(k)$ represents the s-dimensioned vector of moves for all manipulated variables at a given time interval, O_0 is the vector of initial condition, and $d(k)$ is a vector of unmodelled factors. For $r = s = 1$ this equation reduces to model Eq. (8.102).

We can describe a multivariable system dynamic matrix A, which is composed of blocks of dimension $P \times N$ of step response coefficient matrices as in Eq. (8.105) relating the ith output to the jth input as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{bmatrix} \tag{8.113}$$

where elements from matrices a_l have been regrouped accordingly. The matrix A_{ij} contains all the ij coefficients in matrices $a_l = l = 1$ to M arranged as in Eq. (8.105).

The corresponding vector of moves is

$$X\left(\bar{k}\right)=\left[X_1\left(\bar{k}\right)^T\ X_2\left(\bar{k}\right)^T\cdots X_s\left(\bar{k}\right)^T\right]^T \tag{8.114}$$

and the output projection vector becomes:

$$e\left(\bar{k}+1\right)=\left[e_1\left(\bar{k}+1\right)^T\ e_2\left(\bar{k}+1\right)^T\cdots e_r\left(\bar{k}+1\right)^T\right]^T \tag{8.115}$$

Therefore, Eq. (8.110) is equally valid for multivariable systems.

Summary

- Uncertainty is a part of many chemical engineering processes. Fluidized bed, catalytic reactor, composition of available raw material, etc. are very common examples. We have discussed the formulation method of the optimization problem for stochastic method. Stochastic linear and stochastic nonlinear methods are discussed here. Basic knowledge of probability is required to understand the stochastic optimization. Besides this, in the chemical industry, we need to optimize more than one objective functions. Lexicographic method, Linear weighted sum method, evolutionary algorithm and utopia tracking approach can be used to solve multi-objective optimization problem. Controlling different processes is a major task in the chemical process plant. Some advanced control system has been developed based on optimization theories. The error between the measured value and estimated value is minimized.

Review Questions

8.1 Give some examples of uncertainties in chemical industry.

8.2 What is meant by expected value of any random variable?

8.3 When a chemical process is called stochastic process? Give some examples.

8.4 The effluent stream from a wastewater process is monitored to make sure that two process variables, the biological oxidation demand (BOD) and the solid content, meet specifications.

Sample no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
BOD (mg/l)	17.7	23.6	13.2	25.2	13.1	27.8	29.8	14.3	26.0	23.2	22.8	20.4	17.5	18.4	16.8	13.8	19.4	24.7	16.8	27.6
Solids (mg/l)	1380	1458	1322	1448	1334	1485	1503	1341	1448	1426	1417	1384	1380	1396	1345	1349	1398	1426	1361	1476

Find the average BOD and solid content of the effluent.

8.5 Develop the optimization problem of a fluid flow system with two objectives, minimization of installation cost as well as minimization of pumping cost.

8.6 What do you mean by utopia point? What is the significance of that point?

8.7 Write down the algorithm for solving a multi-objective optimization problem using Lexicographic method.

8.8 Why Evolutionary Multio-bjective Optimization is superior to conventional multi-objective optimization?

8.9 Develop optimization problem for controlling a batch reactor with a reaction

$$A + B \xrightarrow{k_1} C + D$$

where C is the desired product.

8.10 What are the advantages of MPC over PID control? Explain your answer considering both economic and technical point.

8.11 Write down the algorithm of a regulatory control system.

References

Al-Othman, W. B. E., Lababidi, H. M. S., Alatiqi, I. M., Al-Shayji, K. *Supply Chain Optimization of Petroleum Organization Under Uncertainty in Market Demands and Prices,* European Journal of Operational Research, 189(2008): 822–40.

Bäck, T., Hammel, U. and Schwefel, H. P. 1997. *Evolutionary Computation: Comments on the History and Current State.* 'IEEE Transactions on Evolutionary Computation', 1(1): 3–17.

Benavides, P. T. and Diwekar U. *Optimal Control of Biodiesel Production in a Batch Reactor Part I: 'Deterministic Control',* Fuel, 94(2012): 211–17.

Charnes, A. and Cooper, W. W. 1959. *Chance Constrained Programming, Management Science,* vol. 6, pp. 73–79.

Cutler, C. R. and B. L. Ramaker. 1980. *Dynamic Matrix Control-a Computer Control Algorithm.* Proc. Joint Automatic Control Conj., San Francisco , CA, Paper WP5–B.

Darby, M. L., Nikolaou, M., Jones J., Nicholson D., RTO: *An Overview and Assessment of Current Practice,* 'Journal of Process Control', 21(2011): 874–84.

Darby, M. L., Harmse, M., Nikolaou, M. 2009. *MPC: Current Practice and Challenges,* Istanbul: ADCHEM.

Deb, K. 2001. *Multi-Objective Optimization using Evolutionary Algorithms,* UK: John Wiley and Sons.

DeCoursey W. J. 2003. *Statistics and Probability for Engineering Applications;* Newnes.

Dehling, H. G., Hoffmann, A. C. and Stuut, H. W. 1999. *Stochastic Models for Transport in a Fluidized Bed.* SIAM J. Appl. Math. 60, pp. 337–58.

De Souza, G., Odloak D., Zanin A. C. *Real Time Optimization (RTO) with Model Predictive Control (MPC);* 'Computers and Chemical Engineering', 34(2010): 1999–2006.

Diwekar, U. 2003. *Introduction to Applied Optimization. 2nd ed.* Boston, MA: Kluwer Academic Publishers.

Diwekar U. 1995. *Batch Distillation: Simulation, Optimal Design and Control.* 2*nd ed.* Washington D. C.: Taylor and Francis.

Dixon, L. C. W. and Szegö, G. P. 1975. *Towards Global Optimisation,* 1–2, North-Holland, Amsterdam, The Netherlands.

Ehrgott, M. 2005. *Multicriteria Optimization,* Second edn., Springer, Berlin.

Fettaka S., Thibault J. and Gupta Y. *Design of Shell and Tube Heat Exchangers using Multi-objective Optimization,* 'International Journal of Heat and Mass Transfer', 60(2013): 343–54.

Garcia, C. E. and Morshedi, A. M. 1986. *Quadratic Programming Solution of Dynamic Matrix Control (QDMC),* 'Chemical Engineering Communications', 46: 1–3: 73–87.

Gattu, G., Palavajjhala, S. and Robertson, D. B. 2003. *Are Oil Refineries Ready for Non-linear Control and Optimization?* in: International Symposium on Process Systems Engineering and Control, January, Mumbai.

Henrion, R., LI, P. Möller, A. *et al. Stochastic Optimization for Operating Chemical Processes Under Uncertainty,* ZIB-Report 01–04 (May 2001).

Hou, K., Li, Y., Shen, J. and Hu, S. *Delayed Sampling Approach to Stochastic Programming in Chemical Processes,* 'Computers and Chemical Engineering' 24(2000): 619–24.

Jonsbraten, T. W. 1998. *Optimization Models for Petroleum Field Exploitation.* Ph.D. thesis. Norwegian School of Economics and Business Administration, Norway: Stavanger College.

Kim, I. Y. 2004. *Adaptive Weighted Sum Method for Multi-objective Optimization* 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.

Li, S. J. Wang, H. and Qian F. 2003. *Multi-objective Genetic Algorithm and its Applications in Chemical Engineering,* Comput. Appl. Chem., 20(6): 755–60.

Logist, F. Van Erdeghem, P. Van Impe, J. 2009. *Efficient Deterministic Multiple Objective Optimal Control of (bio)Chemical Processes.* Chem. Eng. Sci., 64: 2527–38.

Lundstrom, P., Lee, J. H. Morari, M. and Skogestad S. 1995. *Limitations of Dynamic Matrix Control,* Computers Chemical Engineering, 19(4): 409–21.

Marler, R. T. Arora, J. S. *The Weighted Sum Method for Multi-objective Optimization: New Insights,* Struct Multidisc Optim., (2010)41: 853–62.

Marler, R. T., Arora, J. S. *Survey of Multi-objective Optimization Methods for Engineering.* Struct Multidiscipl Optim., 26(2004): 369–95.

Marlin, T. E. and Hrymak, A. N. 1997. *Real-time Operations Optimization of Continuous Processes.* In Proceedings of CPC V, AIChE Symposium Series, vol. 93, pp. 156–64.

Miettien, K. 1999. *Nonlinear Multi-objective Optimization,* Kluver Academic Publishers, Boston, London, Dordrecht.

Muske, K. R., Rawlings, J. B. 1993. *Model Predictive Control with Linear Models,* AIChE Journal, 39(2): 262–87.

Noureddini, H., Zhu D. 1997. *Kinetic of Transesterification of Soybean Oil.* J Am Oil Chem Soc, 74: 1457.

Qin, S. J., Badgwell T. A. 2003. *A Survey of Industrial Model Predictive Control Technology.* Control Engineering Practice, vol. 11, pp. 733–64.

Rao, S. S. 2009. *Engineering Optimization: Theory and Practice,* John Wiley and Sons, Inc.

Rico-Ramirez, V., Diwekar, U. 2004. *Stochastic Maximum Principal for Optimal Control Under Uncertainty.* 'Computers Chemical Engineering', 28: 2845.

Sahinidis, N. V. *Optimization Under Uncertainty: State-of-the-art and Opportunities;* 'Computers and Chemical Engineering', 28(2004): 971–83.

Shrikant, A. B., Saraf D. N. *Steady State Identification, Gross Error Detection, and Data Reconciliation for Industrial Process Units,* 'Industrial and Engineering Chemistry Research', 43(2004): 4323–36.

Stadler, W. 1979. *A Survey of Multicriteria Optimization, or the Vector Maximum Problem, Journal of Optimization Theory and Applications,* vol. 29, pp. 1–52.

Tosukhowong, T. Lee, J. Lee, J. and Lu J. *An Introduction to a Dynamic Plantwide Optimization Strategy for an Integrated Plant.* 'Computers and Chemical Engineering', 29(2004): 199–208.

Weifeng, H., Hongye, S., Shengjing, M. and Jian C. 2007. *Multi-objective Optimization of the Industrial Naphtha Catalytic Re-forming Process,* Chin. J. Chem. Eng., 15(1): 75–80.

Ying, C. M., Joseph, B. 1999. *Performance and Stability Analysis of LP-MPC and QP-MPC Cascade Control Systems,* AIChE Journal, 45(7): 1521–34.

Zavala, V. M. 2013. *Real-Time Optimization Strategies for Building Systems, Industrial and Engineering Chemistry Research,* Ind. Eng. Chem. Res., 52: 3137–50.

Zavala, V. M., Flores-Tlacuahuac, A. *Stability of Multi-objective Predictive Control: A Utopia-tracking Approach,* Automatica, 48(2012): 2627–32.